J Psycholinguist Res (2012) 41:475–486
DOI 10.1007/s10936-012-9204-8

Do Irrelevant Sounds Impair the Maintenance of All Characteristics of Speech in Memory?

D. Gabriel · E. Gaudrain · G. Lebrun-Guillaud · F. Sheppard · I. M. Tomescu · A. Schnider

Published online: 13 March 2012
© Springer Science+Business Media, LLC 2012

Abstract Several studies have shown that maintaining in memory some attributes of speech, such as the content or pitch of an interlocutor's message, is markedly reduced in the presence of background sounds made of spectrotemporal variations. However, experimental paradigms showing this interference have only focused on one attribute of speech at a time, and thus differ from real-life situations in which several attributes have to be memorized and maintained simultaneously. It is possible that the interference is even greater in such a case and can occur for a broader range of background sounds. We developed a paradigm in which participants had to maintain the content, pitch and speaker size of auditorily presented speech information and used various auditory distractors to generate interference. We found that only distractors with spectrotemporal variations impaired the detection, which shows that similar interference mechanisms occur whether there are one or more speech attributes to maintain in memory. A high percentage of false alarms was observed with these distractors, suggesting that spectrotemporal variations not only weaken but also modify the information maintained in memory. Lastly, we found that participants were unaware of the interference. These results are similar to those observed in the visual modality.

D. Gabriel (✉) · I. M. Tomescu · A. Schnider
Division of Neurorehabilitation, Department of Clinical Neurosciences, Geneva University Hospitals, 1211 Geneva 14, Switzerland
e-mail: damiengabriel@yahoo.fr

E. Gaudrain
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

G. Lebrun-Guillaud
CNRS-UMR 5020, Université Claude Bernard Lyon I, Lyon, France

D. Gabriel · F. Sheppard
Clinical Investigation Center, Inserm CIT 808, Besancon University Hospital, 2 place Saint-Jacques, 25030 Besancon, France

I. M. Tomescu
Functional Brain Mapping Laboratory, Department of Fundamental Neuroscience, University Medical Center, CMU, Geneva, Switzerland

Keywords Auditory short-term memory · Speech interference · Change deafness · Auditory illusion

Introduction

One crucial parameter in a conversation between two speakers is the ability to memorize and keep relevant information from the interlocutor's discourse in short-term memory. Performed effortlessly in a quiet environment, these mnesic operations tend to be difficult in noisy situations, not only because an increased listening effort is needed (Surprenant et al. 1999), but also because background noise can blur short-term memory.

To explain this interference, Baddeley's working memory model (Baddeley 1986, 1992, 2000) supports the existence of a temporary, phonological store in which information is maintained for a few seconds but can easily be disrupted by irrelevant sounds. The amount of disruption depends on two variables: the nature of the speech material to be remembered and the acoustic characteristics of the background noise. Indeed, only sounds with changing-state characteristics have been known to impair the maintenance of speech information (Jones et al. 1992; Salamé and Baddeley 1989), regardless of the nature of these irrelevant sounds (music, speech, tones, etc.).

Since speech is made of multiple characteristics varying concomitantly (e.g. content, prosody, etc…), the nature of speech material that has to be remembered is large and not limited to one single element. Nevertheless, most studies have focused only on the content of a message, which is the most important part of information to be stored in memory. The maintenance of the content of a speaker's message has been extensively investigated through serial recall tasks requiring participants to maintain in memory a list of letters or words for a certain length of time (e.g. Colle and Welsch 1976). The recall of serial sequences of verbal items presented auditorily or visually is severely reduced when these sequences are followed by irrelevant sounds, even if participants are instructed to ignore these sounds. This phenomenon is named *irrelevant sound effect* (Jones and Macken 1993; Jones et al. 2004).

The maintenance of pitch information has been investigated through other paradigms, usually comparison tasks requiring participants to compare whether the pitch of a speech sound is the same as the pitch of another sound heard a few seconds earlier. Evaluating whether two speech sounds are of similar pitch has been shown to be severely impaired when some interfering sounds of varying pitch are played at intervals (Deutsch 1972; Semal et al. 1996). A parallel may possibly be drawn between this observation and the effect of phonological similarity in serial recall of word sequences. Salamé and Baddeley (1982) observed an increased amount of disruption when the irrelevant distracting material presented some degree of phonological similarity with the material to be rehearsed (e.g. the word *three* vs. the word *tee*).

Although evidence collected from these different paradigms has led to some conclusions on how short-term memory maintains speech information, several questions remain unanswered. Since the investigations mentioned above have used distinct paradigms and have restricted their investigation to one characteristic of speech, the experimental situation differs from real-life conversations in which several parameters of speech are subjected to change and have to be maintained in memory. During a conversation it is hard to know which parameter of speech to focus on; as a result memorizing and maintaining information seem much more challenging. A consequence of this increased difficulty could be that in natural conversations, changing-state characteristics may not be a required parameter to disturb short-term memory.

The aim of this experiment was to further explore the mnesic mechanisms triggered by the presence of irrelevant sounds. Our primary objective was to reveal whether features of speech affected by sounds with changing-state characteristics were restricted to pitch and content or to other categories. Our hypothesis was that not only all characteristics were affected, but since there were several characteristics to retain in memory, irrelevant, so-called non disturbing noises such as a white noise for example, could also hinder memory. Our secondary objective was to further understand the mechanisms underlying the impairment in maintaining mnesic information. We investigated whether the presence of an irrelevant noise made participants unable to detect speech changes or whether participants misattributed changes. In other words, do irrelevant sounds erase or modify the information relative to speech retained in memory?

To answer these questions, we developed a paradigm in which several characteristics of speech had to be memorized and maintained. This paradigm was a comparison task in which a sequence of five syllables was played, followed by another sequence of five syllables. In one syllable in the second sequence, a change could affect either its content or pitch, as already performed in previous experiments, but also the speaker size. The linguistic content was limited to the phonemic level, since the maintenance of information in the phonological store consists of sub-lexical items (Colle and Welsch 1976; Salamé and Baddeley 1982). Furthermore, this level has been shown to automatically engage speech specific processes (e.g. Dehaene-Lambertz and Pena 2001). Between these two sequences, one of the following irrelevant noises could be played: either one with spectrotemporal variations, or a broadband noise. In addition to the irrelevant noise, we added a condition in which there was almost no delay between the sequences. Since the encoding of the different speech features probably involves different levels of cognitive load in different processing durations, an excessively short time frame between auditory sequences may result in some impairment in detecting changes for some of these features, and would reveal the time course of their encoding. A long silence was used as control condition. It should be noted that the distractor did not overlap with the sequences in order to avoid any energetic masking and to focus on its effect on memory or attention rather than audibility and intelligibility.

In addition to the measurement of change detection performances, we also explored whether or not participants were able to post-evaluate which irrelevant noise induced the most change deafness. In the visual modality, people usually cannot estimate their detection abilities accurately during change blindness experiments, making them blind to their change blindness (Levin 2002; Loussouarn et al. 2011). Similar results in the auditory modality would suggest that the metacognitive processes underlying both sensory modalities share similar properties.

Materials

Participants

This experiment was carried out on 27 participants (6 males, 21 females; mean age = 23.8, SD = 4.1). All of them self-reported having normal hearing and were paid for their participation. Informed consent was obtained from the participants and experimental procedures were approved by The Ethics Committee of the University Hospital of Geneva.

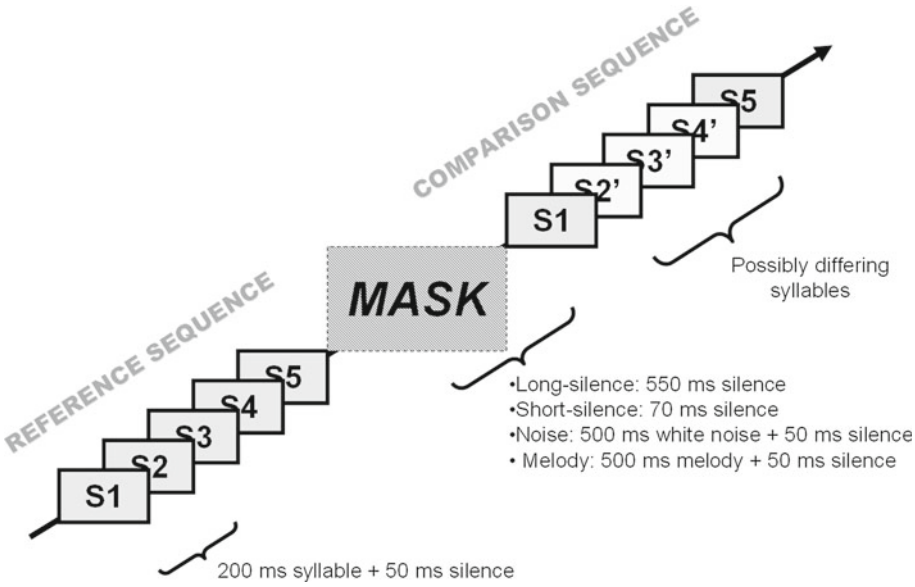

Fig. 1 Schematic illustration of the experimental procedure

Auditory Stimuli

The stimuli were made of sequences of five syllables. A reference sequence was first presented, followed by a mask, followed by a comparison sequence. The comparison sequence differed from the reference sequence in only one syllable, selected at random from the three central syllables (Fig. 1).

Syllables

A set of 60 different consonant-vowel (CV) was recorded from a single male speaker, as described by Ives et al. (2005). The syllables were carefully trimmed to 200 ms, preserving the natural onset and offset. The fundamental frequency (F0) and spectral envelope ratio (SER) were manipulated using STRAIGHT (Kawahara and Irino 2004). A change in F0 induces a change in pitch, while a change in SER affects a change in vocal tract length, and is perceived as a change in the size of the speaker (Smith and Patterson 2005). An SER of 1 corresponds to the original speaker. An SER > 1 corresponds to a spectrum expanded toward the high frequencies, and is related to a shorter vocal tract. On the other hand, an SER < 1 corresponds to a spectrum contracted towards the low frequencies, i.e. a longer vocal tract. All the syllables were equalized to the same intensity.

Sequences

Five different syllables were randomly selected and concatenated to form the reference sequence, with silences of 50 ms after each syllable. Five intervals were randomly chosen from the musical scale (C, D, E, F, G, A and B) and assigned to the syllables. For each sequence, the F0 corresponding to C was randomly chosen between 127 and 202 Hz, so that

the random melodies would not all be tuned on the same key. The SER of each syllable was randomly chosen between 0.95 and 1.20, which roughly corresponds to speakers from 140 to 175 cm tall. The comparison sequence differed from the reference sequence in only one syllable, randomly chosen from the second, third and fourth. The syllable could differ in Pitch (F0), Speaker Size (SER) or Syllable type (consonant). The F0 difference was an increase of 30 % (4.5 semitones). The SER difference was a decrease of 9 % (making the speaker 10 % taller). The syllable change was a consonant substitution: /b/ was replaced by /p/ and vice versa, /g/ by /k/, and /m/ by /n/. The vowels always remained the same, i.e. one of the five canonical vowels /a, e, i, o, u/.

Masks

Three different types of masks were used and compared to a reference condition. The reference was a silence of 550 ms and was thus called Long-silence. The first mask, called Short-silence, was a silence of 70 ms, which is only slightly longer than the interval between two syllables within a sequence. The small time lapse was used to mark the beginning of the comparison sequence and this condition can be considered as an absence of mask. The second mask, named Noise, was a burst of white noise (attenuation of 6 dB/octave) of 500 ms preceded by 50 ms silence. The third mask, called Melody, was a random pure tone melody composed of five different notes. Each note had a duration of 100 ms and its frequency was randomly chosen between 220 and 440 Hz, by steps of 1 semitone. To avoid clicks, 10 ms ramps were applied at the onset and offset of the notes. Thus the random melody was 500 ms, and was also preceded by 50 ms silence.

Stimuli

All changes were combined to each mask type and to the reference, forming 12 different conditions. For each mask, one control condition in which the comparison sequence was identical to the reference sequence was also imposed. In each of these 16 conditions, 10 different stimuli were generated for each of the three possible positions for the change within the sequence, yielding a total of 480 stimuli.

Apparatus

Stimuli were presented diotically via headphones (Sennheiser HD 595) and the intensity was set to 60 dB SPL.

Procedure

Participants were sitting in a sound-attenuating room, and had to press keys 1 (same) or 2 (different) on a keyboard, according to their perception of change. Trials were arranged in 21 different blocks, with a possible short break between each of them. In 1/4 of the trials the comparison sequence was identical to the reference sequence, i.e. the correct answer was "same". All possible combinations of differences and masks were presented in each block. There were thus 48 trials in each block, and the order in which they were presented was random. After each trial, participants pressed a key to start the next sequence 500 ms later.

The experiment lasted approximately 90 min in total. A questionnaire made up of 4 questions was given to the participants at the end of the experiment. They were asked to write down which of the masks (including silence) seemed to be the most disturbing, which was

the least disturbing, which of three changes was the easiest to detect, and which one was the most difficult to point out.

Results

The percentage of correctly detected changes was first analyzed through a two-way repeated measures ANOVA with type of change (syllable, speaker, F0) and type of mask (Melody, Noise, Short-silence, Long-Silence) as factors. The analysis revealed a main effect of type of change ($F(2.52) = 150.47$, $p < 0.0001$), with variations of pitch being the most easily detected features ($p < 0.0001$) (see Table 1). A significant effect of the various masks was also found ($F(7.78) = 7.12$, $p < 0.001$), with better performances with the melodic mask compared to the white noise ($p < 0.01$) and with silence ($p < 0.001$). A subsequent one-way repeated measures ANOVA performed on false alarms nevertheless showed that false alarms also differed significantly among the various masks ($F(3.78) = 11.78$, $p < 0.0001$). The melodic mask induced the highest percentage of false alarms compared to the others ($p < 0.0001$ with Noise and Long-silence; $p < 0.05$ with Short-silence).

These results clearly demonstrate that the higher percentage of correct answers observed with the melodic mask is closely related to a higher number of false alarms. Signal detection theory was thus applied to evaluate the detectability of the changes precisely. The sensitivity index d' was calculated for each mask and each type of change and correction was applied according to Macmillan and Creelman (2005). Mean sensitivity was significantly above zero for all 12 different conditions (each $p < 0.001$), indicating that participants always performed better than chance.

As the aim of this study was to evaluate whether change detection performances were modified by the presentation of a mask rather than to measure these performances for the three different types of changes, statistical analyses were performed on the differences of

Table 1 Percentages of hits and false alarms, Sensitivity d' and response criterion c for all the different types of changes and all the different masks. For each condition, written values are the mean followed by its standard deviation

	Long-silence	Short-silence	White noise	Melody
Syllable				
% Hits	20.6 (14.8)	26.4 (17.9)	23.2 (16.8)	25.5 (14.6)
d'	0.71 (0.5)	0.82 (0.71)	0.77 (0.56)	0.55 (0.49)
c	1.13 (0.42)	0.97 (0.44)	1.07 (0.42)	0.9 (0.39)
Speaker				
% Hits	23.2 (13.4)	21.9 (14.1)	23.2 (13.9)	23.8 (14.2)
d'	0.78 (0.41)	0.6 (0.51)	0.8 (0.43)	0.44 (0.44)
c	1.09 (0.4)	1.08 (0.44)	1.06 (0.41)	0.95 (0.38)
Pitch				
% Hits	75.1 (16.7)	80 (13.9)	75.1 (16.8)	82.7 (12.2)
d'	2.29 (0.84)	2.34 (0.82)	3.29 (0.76)	2.21 (0.81)
c	0.34 (0.36)	0.17 (0.48)	0.32 (0.46)	0.06 (0.28)
% False alarms	7.2 (6.8)	9.8 (9.6)	7.8 (6.8)	12.3 (8.9)

sensitivity between the various masks and Long-silence[1]. Long-silence was chosen as a reference because previous studies have shown that a silence between two auditory sequences does not reduce change detection (Demany et al. 2008; Pavani and Turatto 2008). The decrease in sensitivity, or $\Delta d'$ was calculated by subtracting the d' for a mask and a type of change from the d' for Long-silence and the matched type of change. This decrease was expected to reflect the degree of difficulty. A negative value for $\Delta d'$ represents an improvement in sensitivity.

A two-way repeated measures ANOVA with the following factors: type of change (syllable, speaker, F0) and type of mask (Melody, Noise, Short-silence) was performed to investigate the decrease in change detection sensitivity. As shown in Fig. 2, mask type was found to have an effect ($F(2.52)=5.03$, $p < 0.05$). The detectability of changes was significantly more reduced with the Melodic mask than with other masks ($p < 0.01$ with Noise; $p < 0.05$ with Short-silence). A significant interaction between the type of mask and the type of change was also found ($F(4.104)=2.57$, $p < 0.05$). Detectability of all types of changes was significantly more reduced with the melodic mask (syllable: $p < 0.001$ compared to noise and to short-silence; speaker: $p < 0.001$ compared to noise; $p < 0.05$ compared to Short silence; F0: $p < 0.05$ compared to noise; $p < 0.01$ compared to Short silence). When a change of speaker was detected, the sensitivity was also more degraded with short silence compared to noise ($p < 0.01$). In the melodic condition, the sensitivity was more impaired for a change of speaker size compared to a change of syllable ($p < 0.01$) and pitch ($p < 0.001$). No difference in sensitivity was found between a difference of syllable and pitch.

Statistical analyses were performed on the response criterion c, which represented the participants' tendency to respond that they had detected a change or not independently of the response correctness. For the same reasons as for the measurement of d', statistical analyses were performed by subtracting the c values of the three different masks from the mask in the long-silence condition. The subtracted value, or Δc, shown in Fig. 3, consequently reflected the difference in strategy, also called response criterion, used by the participants when the different masks were presented and was no longer linked to the number of hits. A positive Δc indicated a more liberal criterion in relation to Long-Silence, causing participants to push the 'change' button more often.

The differences in strategy were analyzed through a two-way repeated measures ANOVA with type of change (syllable, speaker, F0) and type of mask (Melody, Noise, Short-silence) as factors. The analysis revealed a main effect of mask type ($F(2.52)=7.81$, $p < 0.01$). Participants had a significantly more liberal approach with the melodic mask than with noise ($p < 0.001$) or short silence ($p < 0.05$), i.e. they were more likely to say there was a change even if there was none. The significant interaction ($F(4.104)=2.57$, $p < 0.05$) also showed that short-silence induced a more conservative criterion when there was a change in speaker size than on other speech characteristics ($p < 0.01$).

Post-evaluation of participants' subjective feelings about the task showed that they correctly assessed the difficulty of a change since all of them perceived Pitch as the easiest one to detect. 63 % of participants considered that speaker size was the least noticeable change. However, participants' evaluation of the difficulty generated by the various auditory masks was less accurate. Only 26 % of them found that melody was the most disturbing mask (Long-Silence: 0 % of participants; Noise: 41 % and Short-silence: 33 %). Nevertheless,

[1] It is also the case that, given the design of the experiment, the false-alarm rate has to be common to the three Types of change. Consequently, the standard calculus for d' could still contain some bias component specific to the subject and the Type of change. However, it seems reasonable to assume that this Type of change specific bias is somewhat independent of the mask. Using the difference of two d' values cancels out this specific bias, and thus also provides a more bias-free sensitivity measure.

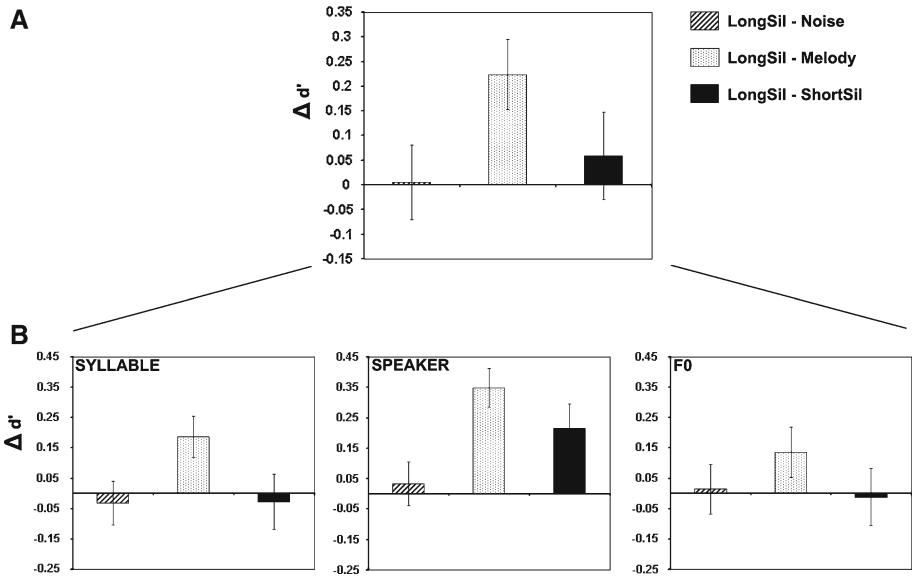

Fig. 2 Decrease in sensitivity relative to the Long-Silence condition. **a** Mean decrease in sensitivity. **b** Decrease in sensitivity for the three different types of changes. A positive value reflects a reduced detectability of changes for the mask compared to Long Silence, and thus can be interpreted as change deafness. *Error bars* indicate the standard error

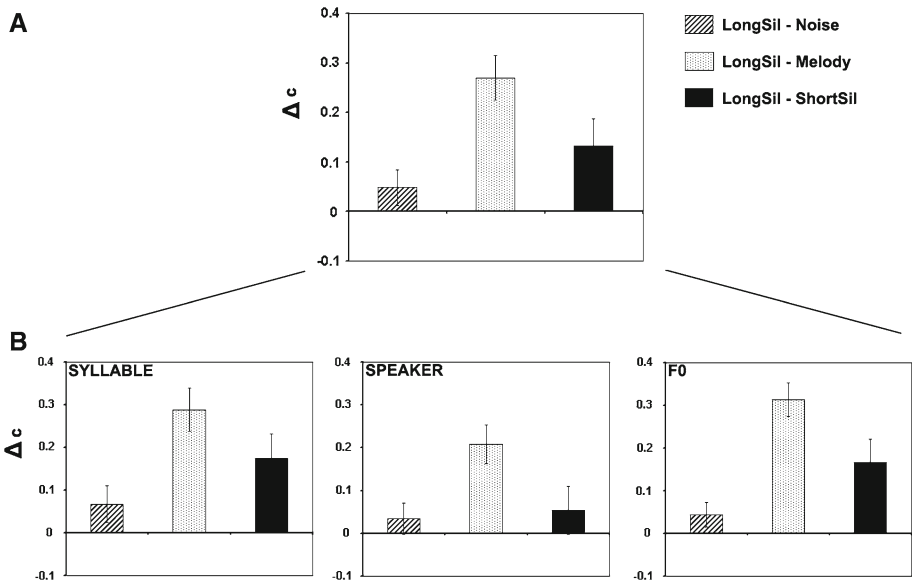

Fig. 3 Difference of response criterion between Long-Silence and other masks. **a** Mean difference of response criterion. **b** Difference of response criterion for the three different types of changes. Note that all differences are positive, suggesting that participants used a more liberal criterion with the various masks than with Long Silence. The more positive the difference, the more liberal was participants' strategy, leading them to give more "change" than "no-change" responses, independently from the response correctness. *Error bars* indicate the standard error

8 % of participants considered it to be the least disturbing one (Long-Silence: 70 %; Noise: 11 %; Short-silence; 11 %).

Discussion

The main result of this experiment is that the deleterious effect of irrelevant sounds appears to be equivalent regardless of how many speech characteristics must be kept in memory. Among the different distractors used, only irrelevant sounds presenting spectrotemporal variations altered the memorization of all speech attributes, whereas broadband noise did not. The discriminability of changes, known as d' according to signal detection theory, was significantly more reduced with the melodic distractor than with others. This is reminiscent of several recall studies showing that tasks involving rehearsal were more disrupted by sounds with changing-state characteristics than by steady-state ones (Banbury and Berry 1998; Salamé and Baddeley 1990). Among the hypotheses to explain these interferences, the feature model (Nairne 1990; see Neath 2000) and the primacy model (Page and Norris 1998, 2003) have suggested that sounds made of spectrotemporal variations require more attentional resources than stationary or repeated items. This could be an extension of the effect of phonological similarity in the phonological model where the degree of interference is related to the degree of similarity between the irrelevant speech stimuli and the items to be remembered. Because of this phonological similarity, "there are fewer discriminating features between items, leading to impaired retrieval and poorer recall" (Baddeley 1992). In our experiment, the structure of the melodic distractor was more similar to the speech sequences than the broadband noise and could have engaged at least a partly similar mechanism.

An additional and intriguing result is that the sound with spectrotemporal variations led participants to modify their response strategy. A more negative response criterion value was observed, which means that participants more often reported that a change had occurred even when this was not the case. This explains why more changes were recorded with the melodic distractor, whether there was a change (hits) or not (false alarms). The increased number of false alarms suggests that participants specifically experienced some "illusions" of change when the melodic distractor was played. To our knowledge, such illusions have not been reported before. Specifically, no information on the percentage of false-alarms and on the response criterion was provided on previous speech comparison experiments using irrelevant sounds with spectrotemporal variations (Semal et al. 1996). Consequently, the presence or absence of the same phenomenon cannot be assessed. A potential explanation may be hypothesized by drawing a parallel with serial recall experiments in which phonological errors, like the phoneme exchange errors that occur in spoonerisms, have frequently been reported (e.g. Morton 1964; Crowder 1978; Ellis 1980). According to the working memory model, these errors arise in the phonological store, on the basis of similarity. Since the information is stored as sub-lexical units in the phonological store, meaningless syllables are more subjected to disruption than concrete words by speech-like irrelevant sounds. The high-level of confusion may have modified the contour (among others) of the original sequence, and negatively impacted the subvocal rehearsal of syllables in a way that could have led to a false perception of a change.

The hypothesis of an interference generated only by sounds with spectrotemporal variations fits well with our results obtained with the white noise. The discriminability of the changes was not reduced by the presence of this irrelevant sound and the number of speech characteristics that had to be remembered thus did not alter the impact of this distractor. This absence of alteration is in accordance with recall experiments showing that bursts of

white noise are significantly less disruptive than spoken words (Salamé and Baddeley 1982; Ellermeier and Hellbrück 1998). Recent studies comparing multiple simultaneous non-speech stimuli also failed to show any difference in the detection of a change whether there is a white noise or silence (Pavani and Turatto 2008). Compared to the melodic distractor, white noise is a random and featureless signal which cannot be precisely stored in memory, except in terms of loudness. Played during the rehearsal of auditory information, it cannot interfere with the contour of the reference sequence and thus cannot lead to any impairment and therefore to any illusion of a change.

In addition to the main effect of distractors with spectrotemporal variations on the maintenance of all speech characteristics, specific effects of some distractors were found for some precise characteristics. Although these data must be taken cautiously since the design of the experiment did not match the difficulty in detecting changes to all speech characteristics, preliminary results provide several interesting topics for future research. The first result is that changes in syllables and changes in pitch were equally affected by distractors. This suggests that verbal content of speech was not as deeply ingrained in memory as its indexical attributes, or at least no more preserved from disturbance (Semal et al. 1996; Stern et al. 2007). As mentioned before, this observation may rely on the absence of meaning in speech sequences to remember, so the content could then be less well preserved. Further investigations using concrete words as stimuli are necessary to assess whether the impairment of speech content depends on the meaning of these stimuli or not.

An additional condition in which the detactibility of changes differed between speech characteristics was when there was almost no pause at all between the sequences. In this case, the discriminability of a change in content or pitch remained unaffected, confirming previous results (Demany et al. 2008), but was significantly reduced for a change of speaker. A potential explanation for this result is that encoding a speaker's voice in memory takes longer than encoding the variations of pitch or content. Whereas it is important to follow some rapid variations of F0, e.g. in prosody, and to encode the content of a sentence, a speaker usually remains the same in a conversation and the listeners exploit this difference of variation rate when establishing speaker identity (Gaudrain et al. 2009). As a consequence, it may not need to be encoded as quickly as the other parameters. The high value of the response criterion for a speaker's voice in the short-silence condition may then be linked to an incomplete encoding of this information, leading to a false perception of a change of speaker. This effect is consistent with the common finding that a noise-reduction algorithm does not improve intelligibility, contrary to self-evaluation by listeners (Hu and Loizou 2007; Sarampalis et al. 2009).

Our last finding is that participants remained unaware of the deleterious impact induced by an irrelevant sound made of spectrotemporal variations. Indeed, only 26 % of participants considered the melodic distractor as the most confusing in their perception of changes. This result is consistent with observations made in the visual modality, in which participants fail to anticipate or recognize their own difficulty in detecting changes (Loussouarn et al. 2011), a phenomenon called change blindness blindness. Similar metacognitive processes probably occur in both sensory modalities with regards to predicting change occurrence.

In light of this study, it appears that an irrelevant sound with distinct spectrotemporal variations similar to speech or music for example, can impair the detection of changes in all speech dimensions. Such an irrelevant sound is quite common in real-life situations, such as a background conversation or music in a restaurant, for example. Even if listening conditions are intact, and even though we may not feel especially disturbed by this interference, these background noises have a deleterious impact on the retrieval of memorized information.

Acknowledgments The authors wish to thank Gaelle Brunotte for her helpful comments and suggestions. This work was partly supported by the UK Medical Research Council (G9900369).

References

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.

Baddeley, A. D. (1992). Is working memory working? The fifteenth Barlett lecture. *Quarterly Journal of Experimental Psychology, 44*, 1–31.

Baddeley, A. D. (2000). The episodic buffer: A new component to working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89*, 499–517.

Colle, H. A., & Welsch, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 6*, 49–54.

Crowder, R. G. (1978). Mechansims of auditory backward masking in the stimulus suffix effect. *Psychological Review, 85*, 502–524.

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12*, 3155–3158.

Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection: Simple sounds are not memorized better than complex sounds. *Psychological Science, 19*, 85–91.

Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science, 175*, 1020–1022.

Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in "irrelevant speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1406–1414.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*, 624–634.

Gaudrain, E., Li, S., Ban, V. S., & Patterson, R. D. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. In *Proceedings of the 10th annual conference of the international speech communication association (Interspeech 2009)*. Brighton, UK.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America, 122*, 1777–1786.

Ives, D. T., Smith, D. R. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *The Journal of the Acoustical Society of America, 118*, 3816–3822.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing-state. *Quarterly Journal of Experimental Psychology, 44*, 645–669.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 369–381.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 656–674.

Kawahara, H., & Irino, T. (2004). Principles of speech manipulation system STRAIGHT. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–179). Boston: Kluwer Academic.

Levin, D. T. (2002). Change blindness blindness as visual metacognition. *Journal of Consciousness Studies, 9*, 111–130.

Loussouarn, A., Gabriel, D., & Proust, J. (2011). Exploring the informational sources of metaperception: The case of Change Blindness Blindness. *Consciousness & Cognition, 20*(4), 1489–1501.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Morton, J. (1964). A preliminary functional model for language behaviour. *International Audiology, 3*, 215–225.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.

Neath, I. (2000). Modelling the effect of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7*, 403–423.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Page, M. P. A., & Norris, D. (2003). The irrelevant sound effect: What needs modelling and a tentative model. *Quarterly Journal of Experimental Psychology, 56A*, 1289–1300.

Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception & Psychophysics, 70*, 619–629.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150–164.

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology, 41A*, 107–122.

Salamé, P., & Baddeley, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society, 28*, 540–542.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech Language and Hearing Research, 52*, 1230–1240.

Semal, C., Demany, L., Ueda, K., & Hallé, P. A. (1996). Speech versus nonspeech in pitch memory. *The Journal of the Acoustical Society of America, 100*, 1132–1140.

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America, 118*, 3177–3186.

Stern, S. E., Mullennix, J. W., Corneille, O., & Huart, J. (2007). Distortions in the memory of the pitch of speech. *Experimental Psychology, 54*, 148–160.

Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory, 7*, 405–420.